



BAB II

TEKNOLOGI INFORMASI

UNIVERSITAS DARMA PERSADA

BAB II

LANDASAN TEORI

2.1 Kajian Penelitian Terkait

Dalam beberapa tahun terakhir banyak penelitian menggunakan *data mining*. Salah satu metode itu banyak digunakan dalam pengelompokan penjualan produk atau barang dalam mengolah data yang banyak dan kompleks untuk menentukan banyak atau sedikitnya produk terjual.

Kemudian dalam memprediksi jumlah penjualan dengan metode Multi Regresi Polinomial yang digunakan untuk peramalan yang lebih sering dipakai pada beberapa kasus untuk pengembangan *inventory*.

Dari penelitian (Benri & Herlina, 2015) tentang “Clustering Menggunakan Metode *K-Means* Dalam Pengelompokan Penjualan Produk Pada Swalayan” telah menunjukkan bahwa penggunaan *Data Mining* dapat membantu pengelompokan data jumlah penjualan rendah dan tinggi sehingga stok barang tidak menumpuk.

Dari Penelitian (M. Hasim Siregar, 2018) tentang “Klasterisasi Penjualan Alat-alat Bangunan Menggunakan *K-Means*” telah menunjukkan bahwa penggunaan *K-Means* dapat menentukan kelompok barang laris dan tidak laris dari hasil jumlah penjualan. Namun, terdapat juga penelitian lain yang menunjukkan bahwa penggunaan data mining tidak selalu memberikan hasil yang lebih baik.

Penelitian lain yang dilakukan (Fitria & La ode dkk, 2019) tentang “Penerapan Metode Regresi Polinomial Pada Pengembangan Aplikasi *Inventory*” menunjukkan mampu melakukan prediksi dengan baik dalam menentukan arah kurva dan prediksi periode berikutnya.

Dari penelitian (Normah, Siti & dkk, 2021) tentang Penerapan Data Mining Metode K-Means Clustering Untuk Analisis Penjualan Pada Toko Fashion Hijab Banten, dimana dapat disimpulkan dengan mengelompokkan data clustering yang dideskripsikan menentukan baju mana yang laris, kurang laris dan sangat laris.

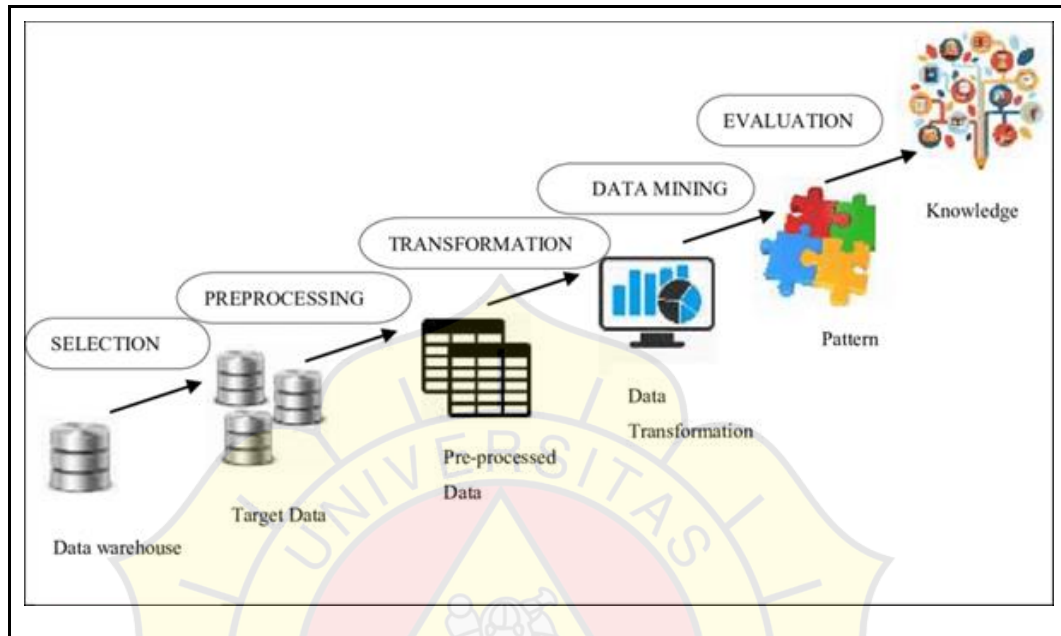
Oleh karena itu, skripsi ini akan mengkaji bagaimana implementasi data mining dengan menggunakan *K-Means* dan Regresi polinomial dapat digunakan untuk memprediksi jumlah penjualan.

2.2 Data Mining

Istilah “penemuan data atau pengetahuan” atau “menemukan pola tersembunyi dalam data” adalah istilah lain dari data mining. Tindakan memeriksa data dari beberapa sudut dan memadatkannya menjadi informasi yang bermakna dikenal sebagai data mining. (Seggall dan lainnya, 2008)

Salah satu teknik dalam melakukan data mining adalah Knowledge Discovery in Database Process (KDD). KDD, sebagaimana didefinisikan oleh Fayyed dkk. (1996), adalah proses penggunaan teknik penambangan data untuk mencari informasi atau pola yang berguna dalam data. Ini memerlukan penerapan algoritma untuk menemukan pola dalam data. Seleksi data, pra-pemrosesan,

transformasi, penambangan, interpretasi, dan penilaian adalah tahapan yang membentuk ringkasan proses KDD Dunham (2003). Berikut penjelasan detail dan contoh metode KDD:



Gambar 2.1 Proses KDD

1. Pemilihan (seleksi) Sebelum dimulainya tahap Penemuan Pengetahuan dalam Basis Data (KDD), ekstraksi data dari sekumpulan data operasional harus diselesaikan. Data pengoperasian dasar disimpan dalam file yang berbeda dari data pilihan yang akan digunakan untuk proses penambangan data.
2. *Pre-processing/ cleaning* Data yang menjadi subjek KDD harus melalui prosedur pembersihan terlebih dahulu sebelum memulai proses data mining. Menghilangkan data yang berlebihan, memeriksa data yang tidak konsisten,

dan memperbaiki kesalahan data seperti kesalahan pencetakan adalah beberapa tugas yang terlibat dalam proses pembersihan. Selain itu juga dilakukan proses pengayaan, yaitu tindakan “memperkaya” data yang sudah ada dengan tambahan data atau informasi relevan dan terkait KDD, seperti tambahan data eksternal yang diperlukan.

3. *Transformation* menjelaskan proses pengkodean. Coding adalah metode transformasi data tertentu agar cocok untuk digunakan dalam data mining. Di KDD, coding merupakan proses kreatif yang sangat bergantung pada jenis atau struktur data yang akan dicari di database
4. *Data mining* adalah praktik penerapan pendekatan atau prosedur tertentu untuk mencari pola atau informasi menarik dalam subkumpulan data. Teknik, metodologi, atau algoritma penambangan data sangat berbeda satu sama lain. Tujuan dan proses umum KDD inilah yang pada akhirnya menentukan apakah pendekatan atau algoritma yang terbaik.

2.3 Clustering

Menurut Tan, 2006 *clustering* adalah tata cara pengorganisasian data menjadi banyak cluster atau kelompok sehingga derajat kemiripan antar data dalam satu cluster paling tinggi dan paling sedikit pada cluster lainnya.

Dalam upaya mengungkap pola tersembunyi pada data yang diteliti, pendekatan clustering harus mampu mengkuantifikasi dirinya sendiri. Untuk mengetahui derajat kemiripan antar objek yang dibandingkan, tersedia beberapa teknik. Jarak

Euclidean Tertimbang adalah salah satunya. Dengan mengetahui nilai masing-masing atribut di dua lokasi, jarak Euclidean menghitung jarak di antara keduanya. Rumus inilah yang digunakan dalam proses penghitungan jarak menggunakan jarak Euclidean:

$$Distance(p, q) = \left(\sum_k^n \mu_k |P_k - q_k|^r \right)^{1/r}$$

Keterangan:

N = Jumlah record data

K= Urutan field data

r= 2

μ_k = Bobot field yang diberikan user

Saat membandingkan dua vektor fitur yang ditentukan berdasarkan peringkat, jarak adalah metode standar yang digunakan untuk menilai seberapa mirip atau berbedanya keduanya. Kemiripan yang lebih besar/dekat antara kedua vektor ditunjukkan jika nilai peringkat akhir lebih kecil.

2.3.1 Karakteristik *Clustering*

Para ahli telah menciptakan berbagai macam teknik pengelompokan. Setiap pendekatan memiliki fitur, kelebihan, dan kekurangan. Ada tiga cara utama untuk membedakan berbagai jenis clustering: kekompakan, keanggotaan dalam cluster,

dan struktur cluster. Teknik clustering dapat diklasifikasikan menjadi dua kategori berdasarkan strukturnya: partisi dan pengelompokan hierarki. Menurut aturan pengelompokan hierarki, satu bagian data dapat dikelompokkan, dua atau lebih kelompok kecil dapat digabungkan menjadi satu kelompok besar, dan seterusnya, hingga semua data dapat diintegrasikan ke dalam satu kelompok. Satu-satunya teknik yang sesuai dengan kategori pengelompokan hierarki adalah pendekatan pengelompokan hierarki.

2.4 K-Means

Teknik pengelompokan yang paling populer, K-Means, digunakan secara luas di banyak domain berbeda karena kesederhanaannya, kemudahan penerapannya, kapasitas untuk mengelompokkan data dalam jumlah besar, dan kompleksitas temporal *linear* $O(nKT)$ Variabel n , K , dan T masing-masing mewakili jumlah dokumen, cluster, dan iterasi.

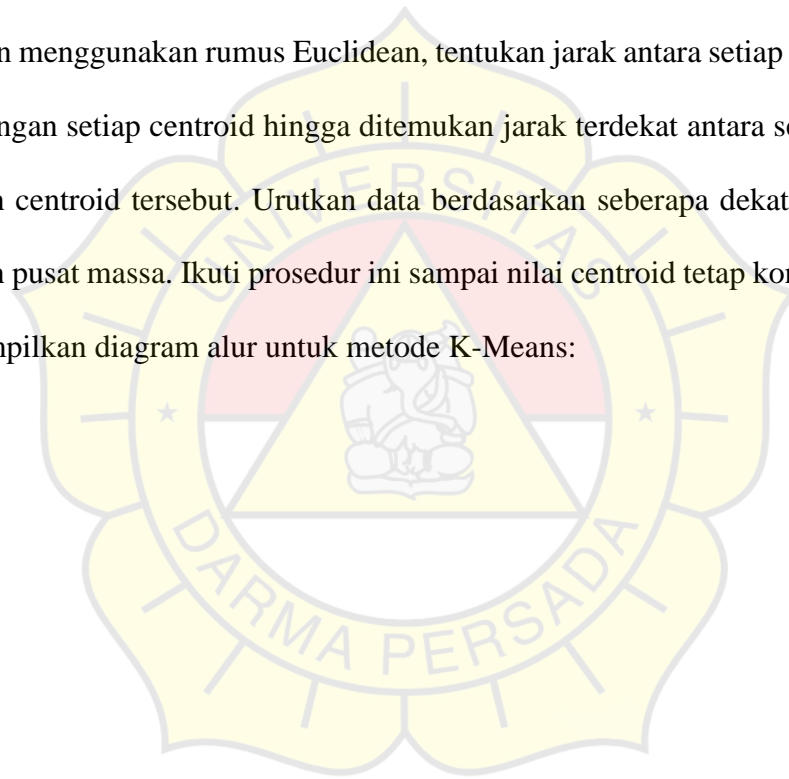
Saat mengorganisasikan data yang sudah ada ke dalam banyak kelompok, pendekatan K-Means dapat membantu dalam mengidentifikasi kualitas yang membedakan data suatu kelompok dari kelompok lainnya sekaligus menjaga kesamaan satu sama lain.

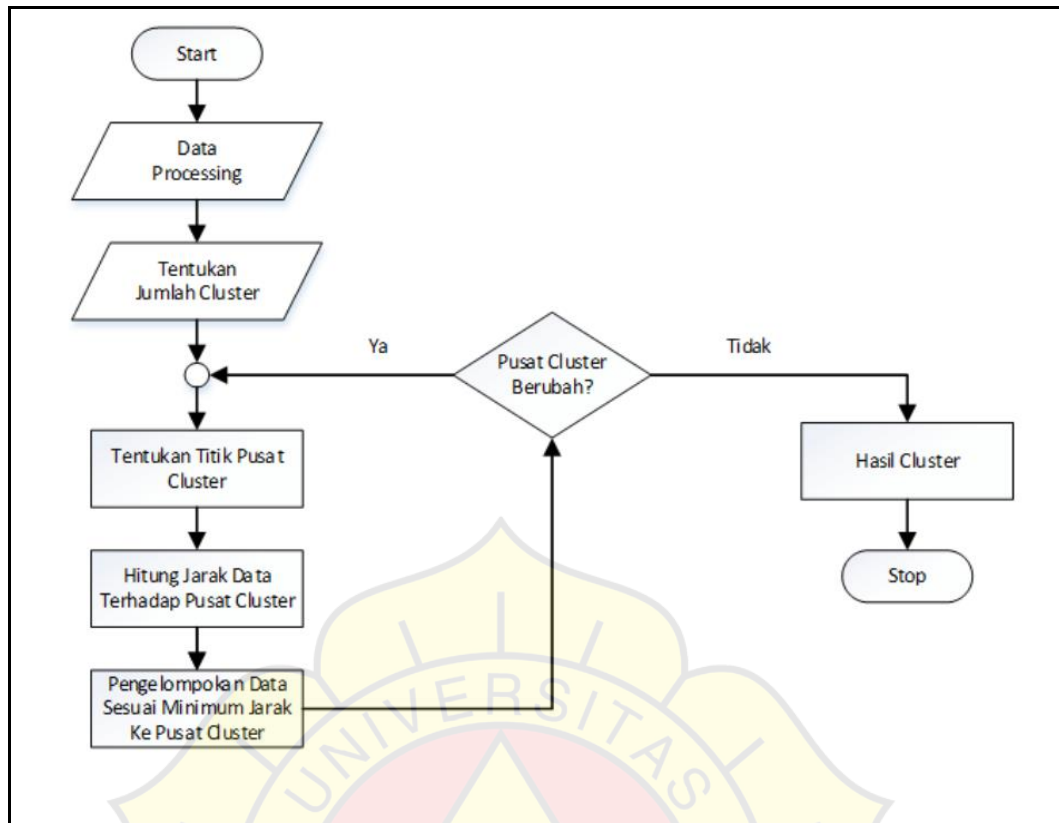
2.4.1 Algoritma K-Means

Pada penelitian menurut (Yulia & Agus, 2016) terkait dengan “penerapan metode k-means dalam mengelompokkan penjualan produk” Hasil dari pengklasifikasian barang yang laku dan tidak laku dapat dilihat oleh pemilik

minimarket yang menggunakan program ini. Pemilik minimarket dapat mencari opsi lain jika barang tertentu tidak laku untuk menjual barang yang tidak laku tersebut.

Algoritma untuk pengelompokan berulang adalah K-Means. Langkah pertama metode K-Means adalah pemilihan K secara acak, dimana K adalah jumlah cluster yang ingin dibangun. Selanjutnya, tetapkan nilai acak ke K. Angka ini akan berfungsi sebagai pusat massa, mean, atau "mean" sementara cluster untuk saat ini. Dengan menggunakan rumus Euclidean, tentukan jarak antara setiap titik data yang ada dengan setiap centroid hingga ditemukan jarak terdekat antara setiap titik data dengan centroid tersebut. Urutkan data berdasarkan seberapa dekat setiap bagian dengan pusat massa. Ikuti prosedur ini sampai nilai centroid tetap konstan. Gambar menampilkan diagram alur untuk metode K-Means:





Gambar 2.2 Algoritma K-Means

Flowchart pada gambar 1, merupakan prosedur penggunaan pendekatan K-Means untuk clustering. Inilah penjelasannya:

Pilih jumlah cluster (“K”) sebagai langkah awal. Pusat cluster “K” dapat ditemukan pada tahap kedua. Ada beberapa metode untuk melakukan hal ini. Hal ini sering dilakukan secara sewenang-wenang. Nomor acak pertama kali ditetapkan ke pusat cluster. Nomor acak digunakan untuk menetapkan nilai awal ke pusat cluster.

Menyelaraskan semua informasi dan item dengan cluster berikutnya adalah fase ketiga. Jarak adalah ukuran seberapa dekat dua benda satu sama lain. Jarak antara data dan pusat cluster juga menentukan seberapa dekat cluster tertentu dengan data. Menghitung jarak antara setiap pusat cluster dan semua data sekarang diperlukan. Sepotong data dikaitkan dengan sebuah cluster berdasarkan jarak terbesar antara nilainya dan cluster tersebut. Dengan rumusan teori jarak Euclidean sebagai berikut, kita dapat menentukan jarak antara setiap pusat cluster dengan seluruh data:

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2}$$

Keterangan:

d = Jarak

j = Banyaknya data

c = Centroid

x = Data Langkah k

Untuk menghitung pusat cluster menggunakan keanggotaan cluster saat ini, lanjutkan ke langkah empat. Semua data dan/atau objek dalam cluster tertentu dirata-ratakan ke pusat cluster. Median cluster adalah pilihan lain jika Anda menginginkannya. Dengan demikian, pengukuran lain di luar mean (rata-rata) juga dapat digunakan.

Mendefinisikan ulang setiap item menggunakan pusat cluster baru adalah fase kelima. Prosedur clustering selesai bila tidak ada perubahan lebih lanjut pada

pusat cluster. Atau, hingga pusat cluster berhenti bergeser, kembali ke langkah ketiga.

2.5 Regresi polinomial

Analisis data melengkung sering dilakukan dengan regresi polinomial, yang digunakan dalam situasi ketika kekuatan variabel independen lebih besar dari 1. Garis yang paling cocok dalam jenis analisis regresi ini bukanlah "garis lurus", melainkan garis lurus. itu selalu berupa "garis kurva" yang secara akurat sesuai dengan titik data.

Ketika beberapa variabel memiliki eksponen dan sejumlah kecil variabel tidak, metode regresi polinomial adalah yang harus digunakan. Selain itu, dimungkinkan untuk memodelkan data yang dapat dipisahkan secara non-linier, yang memberikan fleksibilitas untuk memilih eksponen yang sesuai untuk setiap variabel, selain memberikan kontrol penuh atas berbagai fitur pemodelan yang dapat diakses.

Untuk mengatasi data non-linier, regresi polinomial merupakan perpanjangan dari regresi linier yang menggunakan beberapa variabel. Rumus yang kami gunakan untuk menentukan koefisien regresi β sama dengan rumus yang telah kami bahas sebelumnya pada bagian Regresi Linier Berganda untuk informasi lebih lanjut. Sejak (umam 2018).

$$b = (X^T X)^{-1} X^T y$$

Dalam menghitung nilai koefisien regresi, satu-satunya hal yang membedakan regresi linier polinomial dari regresi linier adalah matriks desain X. Dalam model regresi linier dengan dua variabel independen x dan jumlah total data m, matriks desain dijelaskan sebagai berikut:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$$

Diberikan n observasi, m orde polinomial, dan satu variabel bebas x dalam model regresi polinomial, matriks desain berikut dapat dinyatakan:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ 1 & x_3 & x_3^2 & \dots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix}, b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Model regresi polinomial:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_m x_i^m + \varepsilon_i \quad (i = 1, 2, 3 \dots n)$$

Tidak ada perbedaan antara rumus dan regresi linier berganda dari segi penjelasannya. Untuk menentukan nilai parameter model β , dapat digunakan rumus $b = (X^T X)^{-1} X^T y$. Karena arsitektur matriks yang telah disajikan sebelumnya, kita dapat menghitung koefisien regresi polinomial dengan cara yang sama seperti yang kita lakukan pada regresi linier berganda sebelumnya. Setelah kita memperoleh nilai koefisien regresi, kita dapat menghitung keluaran yang

diharapkan Y dengan menggunakan persamaan $Y = X$. Jika menyangkut orde polinomial, kapasitas model regresi untuk mendeskripsikan data yang rumit meningkat sebanding dengan nilai orde polinomial. Sejak (Umam 2018)

Pada penelitian (Ferryan, Intan & Dkk, 2022) tentang "Peramalan Harga Minyak Mentah Di Indonesia Setelah menghitung angka-angka tersebut melalui model regresi polinomial, kami menemukan bahwa nilai perkiraan terbesar berasal dari minimalisasi kesalahan akar rata-rata kuadrat (RMSE) dan kesalahan persentase absolut maksimum (MAPE) dalam prediksi yang dibuat untuk harga minyak mentah.

Penelitian lain yang dilakukan (Fitria & La ode dkk, 2019) tentang "Penerapan Metode Regresi Polinomial Pada Pengembangan Aplikasi *Inventory*" menunjukkan mampu melakukan prediksi dengan baik dalam menentukan arah kurva dan prediksi periode berikutnya

2.6 Pengertian Persediaan

Aset atau aset lancar yang dimiliki oleh perusahaan yang biasanya menjalankan bisnis melalui penjualan produk atau barang manufaktur disebut inventaris. Berikut beberapa definisi persediaan dari para ahli. Iman Santoso (2010:239) mengartikan persediaan sebagai aset yang dimaksudkan untuk dijual atau diolah lebih lanjut untuk menghasilkan barang jadi, yang kemudian dijual sebagai kegiatan usaha utama.

2.7 Pengertian Suplai

Istilah "persediaan" mengacu pada ketersediaan dan kebutuhan barang. Menyediakan adalah memberikan atau menyediakan barang-barang yang diperlukan (untuk persediaan) (KBBI, 2020). Sebagai kolaborator, pemasok menyediakan komoditas ke bisnis lain. Pembayaran untuk produk seringkali diberikan sesuai jadwal, oleh karena itu pembayaran yang cepat diperlukan sebagai jaminan. Dalam hal ini, untuk menjamin ketersediaan barang yang diperlukan, penyedia sebenarnya harus menganalisis laporan keuangan.

2.8 UML (*Unified Modeling Language*)

2.8.1 Pengertian UML

Istilah UML mengacu pada pendekatan visual untuk membangun sistem berorientasi objek, yang merupakan singkatan dari "*Unified Modeling Language*." Alternatifnya, UML dapat didefinisikan sebagai bahasa yang telah menjadi standar industri untuk desain, dokumentasi, dan visualisasi sistem perangkat lunak. Saat ini, perancangan perangkat lunak pembangunan sebagian besar dilakukan dengan menggunakan UML sebagai bahasa standar. Pemodelan dan pembahasan suatu sistem dengan diagram dilakukan dengan *UML (Unified Modeling Language)*, sebuah bahasa visual.

2.9 Komponen UML

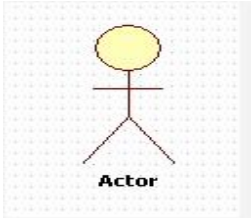
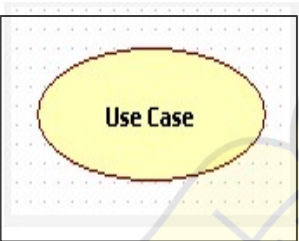
2.9.1 Use Case Diagram

Use Case adalah salah satu instrumen untuk mensimulasikan bagaimana pengguna akan berinteraksi dengan sistem. Tujuan dari penelitian ini adalah menggunakan kasus penggunaan untuk mensimulasikan bagaimana pengguna berinteraksi dengan sistem informasi penjualan. Unified Modeling Language (UML), standar untuk pemodelan visual, pengembangan, dan pendokumentasian sistem yang menghasilkan cetak biru aplikasi, digunakan dalam proses pemodelan kasus penggunaan. Use case adalah kumpulan skenario yang disusun oleh pengguna untuk menyelesaikan suatu tugas (Fowler & Scott, 1993). Kasus penggunaan membantu pemahaman interaksi oleh insinyur perangkat lunak. Aktor-aktor suatu sistem menjalankan beberapa use case, dan sebuah use case itu sendiri dapat memiliki banyak aktor. Ada sejumlah ikatan relasional antara aktor atau use case dan use case, seperti perluasan dan generalisasi.

Use case adalah penjelasan tentang operasi sistem dari sudut pandang pengguna sistem. Kasus penggunaan menentukan data yang akan diproses oleh komponen sistem. Memanfaatkan skenario deskripsi tindakan atau urutan yang menggambarkan apa yang dilakukan pengguna dengan sistem dan sebaliknya use case diimplementasikan. Identifikasi kasus penggunaan sistem, keterlibatan pengguna, dan hubungan fungsi sistem pengguna (Arifin & Hs, 2017).



Berikut adalah beberapa komponen dalam *Use Case Diagram* :

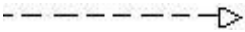
Tabel 2.1 Komponen *Use Case Diagram*.

Komponen Use Case	Penjelasan
	<p>Merupakan bagian yang menjadi ciri individu atau objek (seperti gadget atau sistem lain) yang berkomunikasi dengan sistem.</p>
	<p><i>Use case</i> merupakan gambaran suatu sistem yang dimaksudkan untuk membantu pengguna memahami atau mengapresiasi manfaat dari sistem yang akan dikembangkan.</p>


Ada beberapa relasi yang terdapat dalam *Use Case*, antara lain :

Tabel 2.2 Relasi table *Use Case Diagram*.

Relasi <i>Use Case</i>	Penjelasan
 <p>Relasi <i>Association</i></p>	<p><i>Association</i>, menghubungkan link antara element.</p>
 <p>Relasi <i>Generalization</i></p>	<p><i>Generalization</i>, Suatu komponen mungkin merupakan keistimewaan dari komponen lain; ini dikenal sebagai warisan.</p>

 <p>Relasi <i>Depedency</i></p>	<p><i>Dependency</i>, Suatu komponen saling bergantung dengan komponen lainnya.</p>
--	---

Tabel 2. 3 Tabel *Stereotype*



Relasi / Stereotype	Penjelasan
<p><<include>> *</p> 	<p>Tindakan yang harus dilakukan agar suatu peristiwa dapat terjadi, dimana suatu use case bergantung pada use case yang lain.</p>
<p><<extends>></p>	<p>Contoh perilaku yang hanya efektif dalam keadaan tertentu adalah dengan menyalakan alarm.</p>
<p><<Communicates>></p>	<p>Ditambahkan dalam kasus di mana afiliasi dapat menunjukkan bahwa</p>






	<p>mereka adalah asosiasi komunikasi.</p> <p>Selama asosiasi adalah satu-satunya jenis hubungan yang dapat dibangun antara aktor dan use case, maka ini adalah sebuah pilihan.</p>
--	--

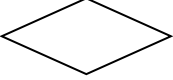

2.9.2 Activity Diagram

Activity Diagram adalah representasi grafis dari tindakan pengguna sistem dari setiap menu sistem. Diagram aktivitas menjelaskan berbagai tugas yang terlibat dalam desain sistem, serta bagaimana setiap kemampuan beroperasi dan diakhiri.

Tabel 2.4 Komponen *Activity Diagram*

<i>Activity Diagram</i>	Penjelasan
 <i>Start State</i>	<i>Start State</i> , sebagai indikasi bahwa prosedur diagram aktivitas telah dimulai.
 <i>State</i>	State, berfungsi menampung <i>event</i> dalam <i>activity diagram</i> .

 <i>Activity</i>	<p><i>Activity</i>, mempunyai peran yang sama dengan pemerintah. Mendorong Kegiatan atau Peristiwa dalam Sistem Proses.</p>
 <i>State Transition</i>	<p><i>State Transition</i>, bertindak untuk mengilustrasikan pada diagram arah atau urutan kejadian atau aktivitas.</p>
 <i>Transition to self</i>	<p><i>Transition to self</i>, membantu mengilustrasikan bagaimana suatu peristiwa berubah sebelum menjadi peristiwa itu sendiri.</p>
 <i>Horizontal Synchronization</i>	<p><i>Horizontal Synchronization</i>, membantu menunjukkan bagaimana suatu peristiwa bertransformasi sebelum peristiwa itu benar-benar terjadi.</p>
 <i>Vertical Synchronization</i>	<p><i>Vertical Synchronization</i>, berfungsi untuk</p>

	mensinkronisasikan 2 cabang <i>event</i> yang posisinya vertikal.
 <i>Decision</i>	<i>Decision</i> , digunakan ketika grafik menampilkan pilihan antara dua keadaan peristiwa.
 End State	End State, sebagai tanda akhir dari <i>activity diagram</i> .

2.9.3 Class Diagram

Dalam Unified Modeling Language (UML), diagram kelas merupakan suatu bentuk diagram yang digunakan untuk menggambarkan kelas dan paket dalam suatu sistem yang akan digunakan di masa mendatang. Konsekuensinya, diagram ini mampu memberikan ringkasan sistem serta hubungan yang ada di dalam sistem.